
Assessing basis risk for longevity transactions – phase 2 presented by Dr Jackie Li and Dr Chong It Tan, IFoA Longevity Basis Risk Working Group

Abstract of the London Discussion

[Institute and Faculty of Actuaries, Sessional Research Event, London, 4 December 2017]

This abstract relates to the following paper: Li, J., Li, J. S. H., Tan, C. I. and Tickle, L. (2018) Assessing basis risk in index-based longevity swap transactions. *Annals of Actuarial Science*. Cambridge University Press, doi: 10.1017/S1748499518000179.

The Chairman (Mr P. H. Simpson, F.I.A.): This session's topic is assessing basis risk for longevity transactions – phase 2. The Life and Longevity Markets Association (LLMA) began publishing indices linked to population mortality statistics in March 2012 with the aim of facilitating the hedging of longevity risk for pension funds and annuity books.

The launch of the LLMA indices was an important milestone towards the longevity market, where risk management can be carried out through transactions that are linked to standardised population level data. In addition to the mortality indices, the LLMA has also produced a significant body of work around possible derivative transactions that could reference mortality indices and offer standardised longevity risk management tools.

However, these building blocks have not proved sufficient to develop a liquid market in longevity and have not led to transactions based on these standardised measures.

We believe that a major obstacle to widespread use of longevity risk management tools as reference population-based mortality indices is the difficulty in quantifying and hence managing longevity basis risk.

In December 2011, the LLMA and the IFoA formed the Longevity Basis Risk Working Group with a remit to investigate how to provide a market friendly means of analysing longevity basis risk.

Phase I of the project was completed by the Cass Business School and Hymans Robertson in December 2014, in which a decision tree framework was developed as a practical guide on how to select a two- population mortality model. It includes the M7 and M5 model and the CAE cohorts model and the characterisation approach.

Phase 2 of the project (which we are discussing here) was commissioned by the IFoA and the LLMA in 2016 and was undertaken by Macquarie University with support from the University of Waterloo and Mercer Australia.

Phase 2 has focused on longevity basis risk in realistic scenarios under practical circumstances.

Our two speakers are Jackie Li, who is an associate professor in Actuarial Science at Macquarie University. He obtained his first PhD in actuarial studies from the University of Melbourne and his second PhD in demography from Macquarie University.

He is a fellow of the Institute of Actuaries of Australia. His research interests are in mortality and longevity modelling, pricing and stochastic reserving methods for general insurance. His research has been published in leading actuarial and demographic journals, including the ASTIN bulletin; Insurance, Mathematics and Economics; The North American Actuarial Journal; The Scandinavian Actuarial Journal; the IFoA's Annals of Actuarial Science; and Population Studies and Demographic Research.

From 2007 to 2014, Jackie worked in Nanyang Business School at Nanyang Technological University in Singapore. He was one of the committee members to the accreditation agreement from the IFoA in 2012. Before he joined academia, he worked as an actuary for a number of years in general insurance and superannuation.

Jackie is accompanied here by his colleague Chong It Tan. Chong is a senior lecturer at Macquarie University and obtained his PhD from Nanyang Technology University. He is a fellow of the Society of Actuaries, a fellow of the Institute of Actuaries Australia and a Chartered Enterprise Risk Analyst. His major research interests are mortality modelling, longevity risk and bonus malus systems. His research has been published in Insurance, Mathematics and Economics and the IFoA's Annals of Actuarial Science.

Dr J. Li: I will talk about our research project, assessing basis risk for longevity transactions – phase 2. This is covered in detail in our report and we highlight some of our major findings at this session.

My research team members are my colleague Professor Leonie Tickle, from Macquarie University, Dr Chong It Tan and Professor Johnny Siu-Hang Li, from the University of Waterloo.

Our focus is on index-based longevity hedging, highlighted in Figure 1.

Consider a pension plan with a certain amount of longevity risk. We call the population underlying the plan the “book population.” Suppose the pension plan sponsor wants to reduce the longevity risk exposure by using a hedging instrument and this instrument is linked to a particular index population or reference population, such as a specific cohort of the England and Wales population.

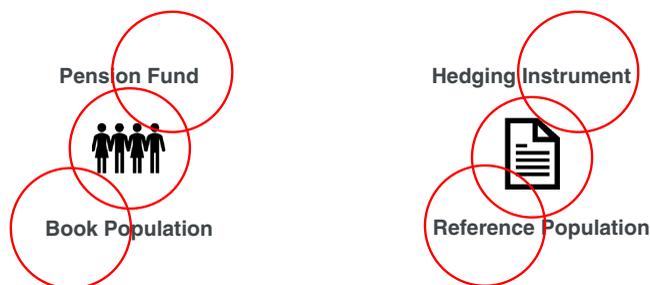


Figure 1. Index-Based longevity hedging

This index-based approach, however, would not give a perfect hedge. First, the book population and the reference population though demographically related are different. They may have different mortality trends.

Second, the number of lives in the pension plan is often much smaller than that in the whole population. The randomness of individual lives could drive the two experiences apart.

Moreover, the plan pays regularly on the survival of the pensioners, while the instrument pays according to its specific design, and the two streams of payments are usually different in their amounts and also timing.

So, we have to find a way to measure the extent of mismatching between the two sides and then work out the actual amount of longevity risk that can be reduced. We call this mismatching Longevity Basis Risk as shown in Figure 2.

We can broadly identify three sources of longevity basis risk: namely, demographic basis risk, caused by demographic or social-economic differences; sampling basis risk, due to random outcomes of individual lives; and structural basis risk, because of the differences in payoff structures.

Phase 1 of this project has provided a framework for modelling demographic basis risk. The current phase 2 takes all the three risk components into account.

Phase 2 has four main objectives as highlighted in Figure 3.

First, to determine the most relevant risk metrics for measuring longevity basis risk and hedge effectiveness. Second, to apply the framework developed in phase 1 to realistic worked examples using appropriate data.

Third, to present a robust quantification of basis risk to third parties like regulators. Fourth, to investigate the limitations of the time series processes.

I want to emphasise that all these objectives are based on the methodology from phase 1 and the main purpose of phase 2 is to apply it but not to develop new methods.

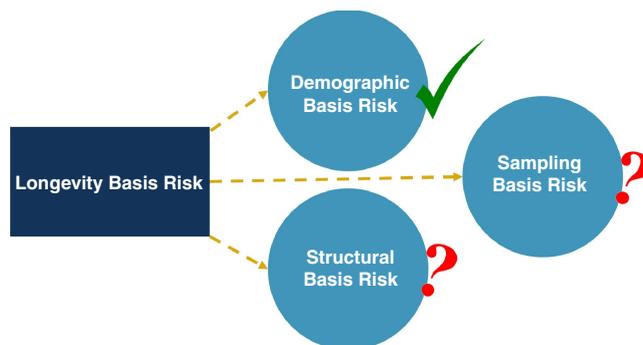


Figure 2. Longevity basis risk

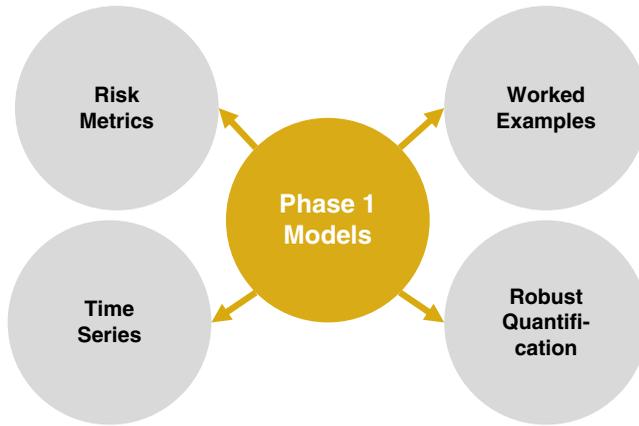


Figure 3. Phase 2 objectives

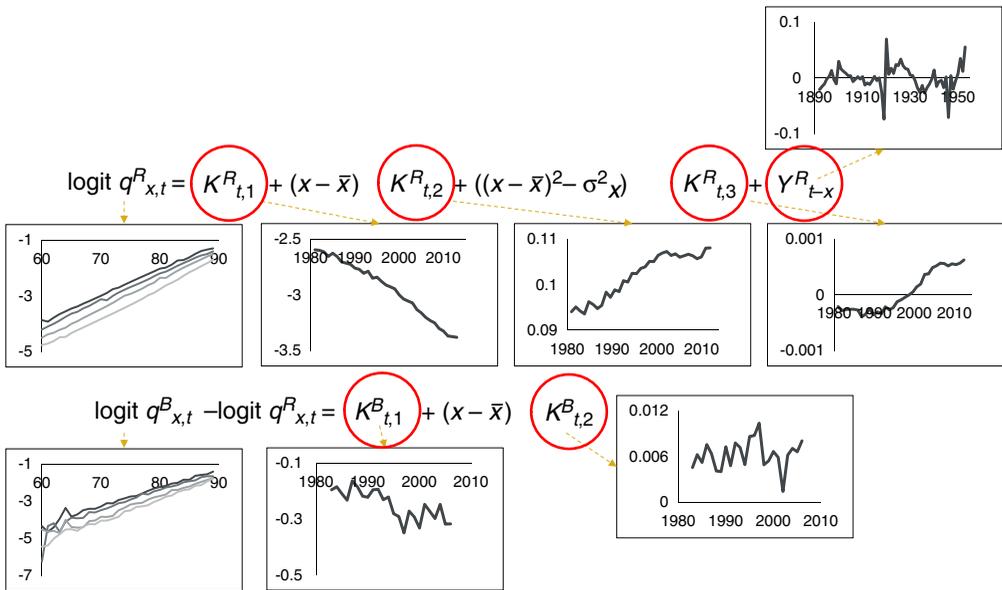


Figure 4. Phase 1 (M7-M5 Model)

Having said that, we still have explored a range of different extensions, and also a number of other models for comparison purposes.

Data are provided by the Continuous Mortality Investigation (CMI), the Office for National Statistics (ONS), Mercer Australia and also the Human Mortality Database (HMD). The first three datasets are taken as the book experience and the last dataset is taken as the reference experience.

Figure 4 gives a quick review of the first method in phase 1 – the M7-M5 model. The first model component is for the reference population: $q^R_{x,t}$ represents the reference mortality rate at age x in year t . Figure 4 shows how the whole mortality curve decreases over time, which reflects the ongoing mortality improvement of the reference population.

On the right-hand side of the equation in Figure 4, $k^R_{t,1}$ refers to the level of the mortality curve, which decreases over time. $K^R_{t,2}$ refers to the slope of the mortality curve, which increases over time: this means the mortality curve gets steeper across time. $K^R_{t,3}$ refers to the curvature of the mortality curve. Its increasing trend means that the mortality curve gets bent more over time. These three k^R s are modelled by a trivariate random walk with drift. The last parameter, γ^R_{t-x} allows for the cohort effect. It is modelled by an ARIMA (1,1,0) process.

The second model component is for the book population: $q^B_{x,t}$ is the book mortality rate. The difference between the logit book rate and the logit reference rate is expressed by another two k^B s, $K^B_{t,1}$ and $k^B_{t,2}$ which are modelled by a VaR(1) process.

The next one, CAE + Cohorts model: in the first component on the left of Figure 5 there is the logit $q^R_{x,t}$ again. On the right, it is a Lee-Carter kind of structure: α^R_x refers to the mortality schedule over age. β^R_x is an age-specific sensitivity measure and k^R_t is the so-called mortality index, which indicates the overall mortality improvement over time. This k^R_t is modelled by a random walk with drift. At the end there is also the cohort parameter, which is modelled by an ARIMA(1,1,0).

In the second model component, the difference between the logit book rate and the logit reference rate is expressed in terms of another Lee-Carter structure with a different α^B_x but the same β^R_x as above. That is why it is called common age effect – CAE. And a different K^B_t , which is modelled by an AR(1) process.

The next option is called the characterisation approach being proposed in phase 1. When the book data size is small, modelling the book data directly may over-estimate the basis risk. So under this characterisation approach, we do not do that. Instead, we further divide the book data into a few categories.

For example: low income, medium income and high income. Then we find some other data that have a large size for modelling and use these as a proxy for each category. Finally, the models are applied

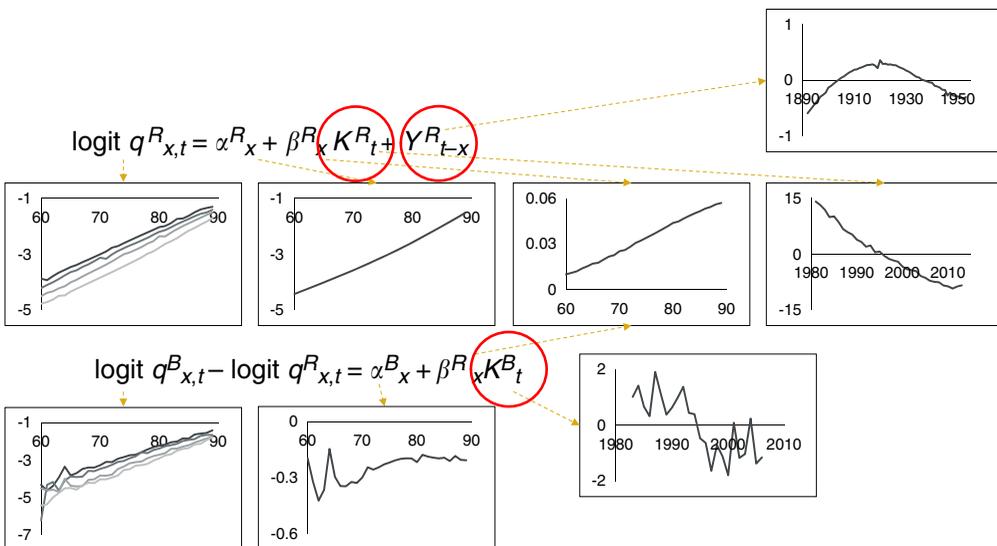


Figure 5. Phase 1 (CAE+Cohorts Model)

and fitted to these proxy data and their simulations are aggregated together and are treated as if they were simulated from the original small book data.

Besides the two models in phase 1, another model that we would like to mention and that we have tested, is the one proposed by Zhou, Li and Tan in 2013, modified for the cohort effect, as in Figure 6. In fact, this model can be seen as a predecessor of the CAE + Cohorts model. In the first component, it has the same $\alpha^R_x, \beta^R_x, k^R_t$ modelled by a random walk with drift and γ^R modelled by an ARIMA (1,1,0) process.

In the second component, the logit book rate itself, but not the difference between the logit book rate and the logit reference rate follows a Lee Carter structure $-\alpha^B_x$, the same β^R_x , and k^B_t . Then the difference between the two k s is modelled by an AR(1) process. And, lastly the same γ^R_{t-x} .

Effectively, the CAE + Cohorts model is a slight mutation of this Zhou, Li and Tan model.

Now we start with a simple hypothetical example of a pension plan. Suppose all pensioners are now aged 65. Each pension pays £1 per year on survival from ages 66 to 90. The pension plan is closed and there are no new members.

A 25-year index-based longevity swap is used to hedge the plan. The size of the swap is calculated from numerical optimisation based on simulated scenarios from the models. The objective of the optimisation is to minimise the longevity risk of the pension liability. Assume that the interest rate is 1% per annum.

Then in the following sensitivity analysis, we also examine a variety of changes to these settings. For example, different plan sizes; multiple cohorts, including older ages, an open pension plan; using several swaps for multiple cohorts; different interest rates; different data fitting periods; alternative

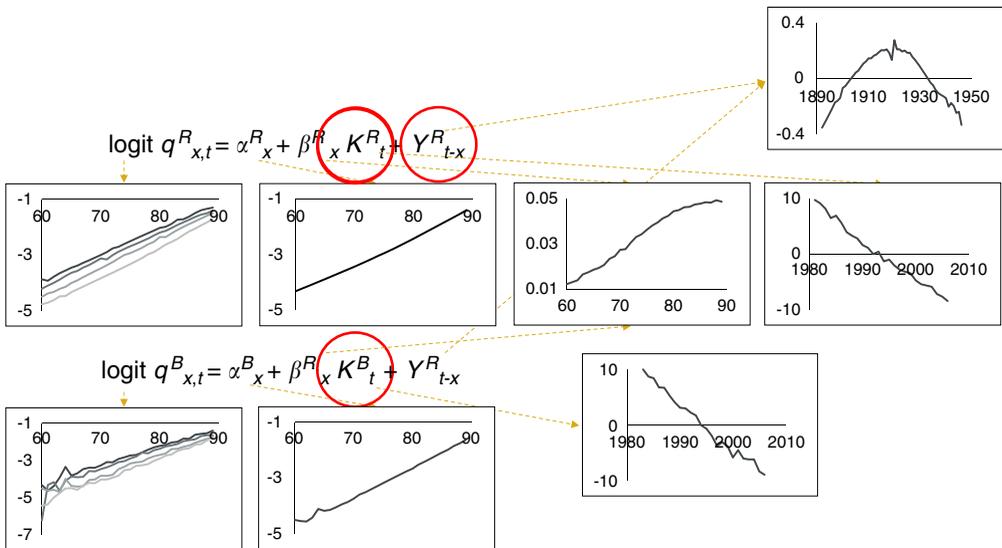


Figure 6. Zhou, Li, Tan (2013) Model + Cohort

simulation methods; additional features like mortality jumps and structural changes; and different time-series processes.

Here I only cover the major ideas. You can find all the details in our report.

We focus on the use of index-based longevity swaps. In current practice, bespoke longevity swaps are by far the most commonly used swaps as hedging instruments. Figure 7 shows how an index-based swap works instead. Consider a pension fund that pays pensions to the pensioners.

These payments depend on the book experience. At the same time if this fund enters into an index-based swap with a certain counterparty and makes fixed regular payments, the fund receives in return floating payments from the counterparty. These payments depend on the reference experience. Clearly, the two sets of cashflows would not be perfectly matched because of demographic basis risk, sampling basis risk and structural basis risk, as mentioned earlier.

It is important, first, to calibrate the swaps carefully in terms of the age buckets used, their maturities and their weights. Then, we estimate the amount of longevity risk that can be reduced by using the swaps.

Regarding the demographic basis risk, Figure 8 shows some simulated examples of how the book mortality rates and the reference mortality rates may move together or deviate from each other over time. We used the CMI data, the HMD data, the CAE + Cohorts model and a single cohort aged 65 in 2014 for this demonstration.

In the first graph on Figure 8, the black lines are for the book population and the grey lines for the reference population. The book mortality is lower than the reference mortality. The mortality rates increase with age for a specific cohort. Moreover, the solid lines refer to the best estimates and the dotted lines refer to the simulated values.

In this simulated scenario, the two populations' future mortality rates follow their expected values closely and move in parallel amongst each other, with everything seeming to go according to plan.

In the second scenario (graph on the right), both populations' future mortality rates deviate from their expected values over time; but since they go hand-in-hand with each other, there is still not much mismatching problem.

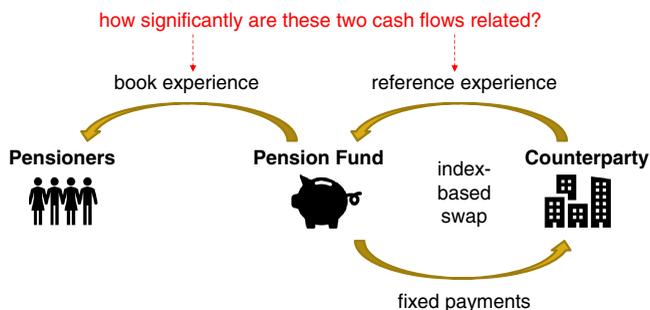


Figure 7. Index-Based longevity swap

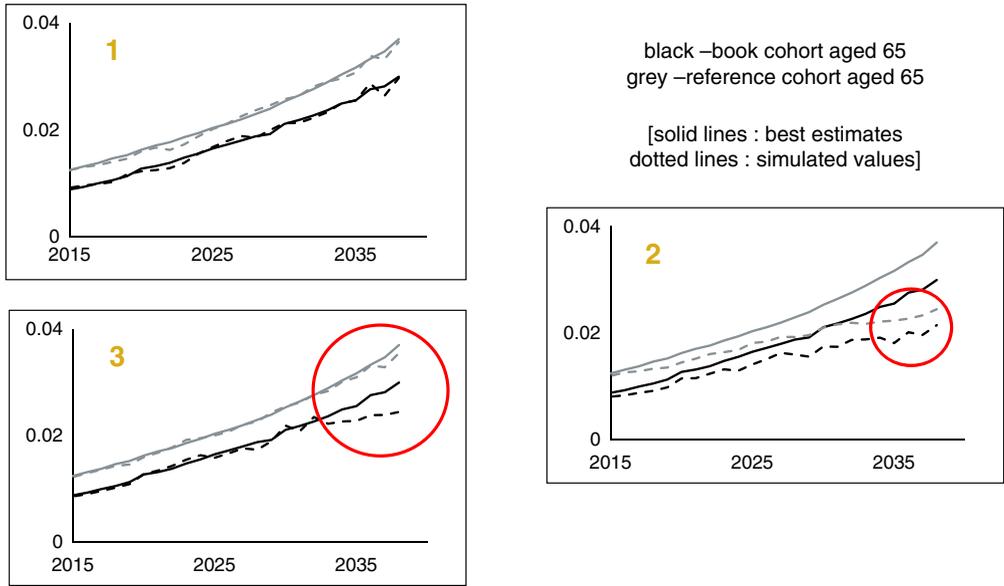


Figure 8. Book vs Reference mortality rates

But in the third scenario (bottom graph), while the reference mortality rates stick closely to the expected values, the book mortality rates diverge from the expected trend and deviate significantly from the reference mortality rates in the later period, being much lower than what was expected. This is the mismatching that we would worry about.

There are certainly many, many other scenarios that could be simulated but they would fall broadly into these three categories.

The question is: what is the probability for each scenario to occur? The last one is the basis risk that we have been talking about. We need one of those two-population mortality projection models mentioned earlier to measure this basis risk.

Using an appropriate model that allows for longevity basis risk, we can measure the effectiveness of an index-based longevity hedge. Figure 9 shows the simulated distributions of the pension liability’s present value using some CMI, ONS and Mercer Australia datasets, before and after taking the hedge.

It is clear that the hedge significantly reduces the simulated variability of the pension liability. In particular, the right tail is very much reduced. This tail may be of much concern in practice and may indicate certain opportunities for capital savings.

To quantify the precise level of longevity risk reduction, we define it as $(1 - \text{risk}_{\text{hedged}} / \text{risk}_{\text{unhedged}})$ in percentage, where $\text{risk}_{\text{unhedged}}$ is the portfolio’s longevity risk before taking the hedge and $\text{risk}_{\text{hedged}}$ is the risk afterwards.

In effect, this metric gives the percentage of the portfolio’s initial longevity risk that is being hedged away. Regarding the risk measures, we considered the common ones, including the variance,

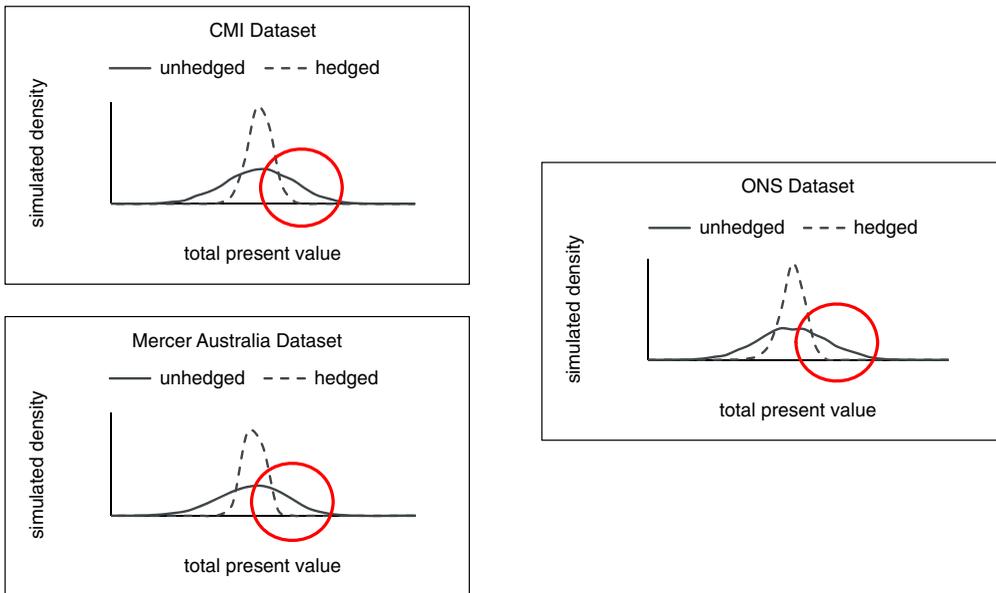


Figure 9. Hedge effectiveness

standard deviation, value at risk and expected shortfall. Although the variance is the one usually adopted in the basis risk literature, this risk measure has a different scale from the other three and it always results in a much higher calculated level of risk reduction than the others. So we focus on the other three risk measures, but not quite the variance.

In addition, the value-at-risk is highly relevant to the solvency capital requirement under Solvency II, and the expected shortfall has recently been suggested to replace the value-at-risk in some banking regulations. So we focus on the standard deviation, the value-at-risk and expected shortfall.

After examining several hundreds of simulated cases from different model settings, different datasets and hedging environments, in which in each case we simulated 5,000 scenarios, we find that for a large pension plan or large annuity portfolio, having around 20,000 lives or more, the risk reduction level is often around 50% to 80%. But for a small plan, or a small portfolio, the risk reduction is usually less than 50%.

These results highlight the impact of sampling basis risk, under which the higher variability in a small book because of individual uncertainty can significantly hamper the effectiveness of an index-based hedge; but note that the precise risk reduction level really depends on the particular circumstances being studied.

We have done an extensive set of sensitivity analysis, and the most important factors that we have found in calculating the risk reduction are:

whether the portfolio size is large;

whether the model is coherent, that is the projected ratio or the expected ratio of the future mortality rates between the two populations at each age converges to a constant in the long-term;

whether the book data is a significant subset of the reference data;

whether the longevity swap or other derivatives like the q-forward are being used;

whether bootstrapping or a simple parametric method is used; and

whether structural mortality changes are incorporated into the modelling process.

Studying the limitations and implications of the time series modelling is one of the main objectives in phase 2. As our book data covers only a very short period of time, many of the sophisticated time series processes cannot be used in our work. But we do test different variations of the time series modelling assumptions and study the corresponding impact.

We find that the most important factor is whether the simulated future variability of the book minus reference component is bounded or not. In fact, despite the obvious differences in the mortality structures between the models, we find that the simulated variability of this model component turns out to be the most important modelling consideration in estimating the risk reduction.

Another important factor is how fast the model could reach coherence – that is, how fast the two populations' future mortality rates would move back in line expectedly. Comparatively, the other correlation assumptions are not so important.

Further research is required when more book data of longer periods and for different portfolios can be collected in the future.

Finally, we summarise the hedging results of our hundreds of simulated cases in a simple qualitative way and also another simple quantitative way.

Industry reports in Australia have used these approaches in the past. Regulators and practitioners may expand these two ideas to develop, say, a prescribed approach of calculating the risk reduction of an index-based longevity hedge, compared to a full internal modelling approach which requires the use of the phase 1 model and other models.

Our findings are potentially useful to pension actuaries and reinsurers. Consider the following very simple example of an actuary managing a pension plan and implementing an index-based hedge. First, if the plan size is large, the risk reduction is often 50% or more based on our simulations.

He would then consider if the two populations are very related. Let's say the answer is six out of ten. Then consider if the two populations' future mortality rates move back in line quickly. Let's say the answer is average, meaning five points out of ten.

Following on he would consider any potential structural changes that affect both populations in about the same way: the answer could be leading to a further four points.

The scores would total 65, which means about 65% of the risk can potentially be hedged away.

Alternatively, a simple linear regression formula can be deduced to summarise all our simulated results. There are ten explanatory variables, including the log portfolio size, how related the two

populations are, the design of the hedging scheme, the interest rate, using swaps or q-forwards, whether M7-M5 is used, whether CAE + Cohorts is used, which simulation method is adopted, whether structural mortality changes are incorporated, and the autoregressive order of the book minus reference component's time series process.

The first five variables refer to the pension plan and the hedging environment. The other five describe the model setting and assumptions.

The signs of the coefficients can tell us how the risk reduction estimate would react to a change in a certain factor. The larger the size, the higher the risk reduction, and so a positive sign. The closer the relationship between the two populations, the higher the risk reduction. We use a categorical variable here to differentiate the relationships. The more approximate the hedging scheme, the lower the risk reduction. Again, we use a categorical variable here. The higher the interest rate, the lower the risk reduction. Using q-forwards decreases the risk reduction.

The next two values indicate that the CAE + Cohorts model tends to give a higher risk reduction estimate. Using a simple parametric method, which does not allow for parameter uncertainty, increases the risk reduction. This highlights the importance of incorporating parameter uncertainty in the simulation. Including structural changes increases the risk reduction because the two populations would move more consistently if they are both influenced significantly by the same factor.

Lastly, the higher the autoregressive order, the slower the convergence of the time series process in the book minus reference component, the more the mismatching and the lower the risk reduction.

This project work, both phases 1 and 2, should encourage a step change in the ability to assess longevity basis risk. Looking forward, it is important to further test more book data, more mortality models and time series processes, and more hedging scenarios to gain a deeper understanding of longevity basis risk and to identify opportunities for potential capital savings.

It is also important to communicate the results properly with different stakeholders, insurers, banks, regulators and clients, on the feasibility and the practical implementation of index-based hedging.

Finally, it would be worthwhile for regulators and the industry to give a standardised and comprehensive list of key factors that drive longevity basis risk, and then practitioners could choose between a prescribed approach, which makes a simple use of these standardised factors, or alternatively an internal modelling approach, which requires necessary expertise to perform the modelling and the computations.

A lot of the details can be found in our final report. The report can be approached in three different ways. If you do not have much time, you can simply go through the graphs and tables, from which you can get some very rough ideas.

For those who want to get more information, you can read the relevant chapters in more detail.

Finally, for those modellers, analysts and academics who have a strong interest in mathematical formulae and programming, you are encouraged to read the two appendices. All the technical details of our work are given in the appendices.

Professor A. J. G. Cairns, F.F.A. (opening the discussion): I am here as director of the actuarial research centre ARC of the IFoA. Although this particular project was commissioned before the ARC was formed in its expanded scope. It is overseeing the commissioned research projects as part of the actuarial profession, so it is very encouraging to see one of these projects come to fruition and to be presented here at our sessional research event. This is our flagship event for the presentation of results, and also discussion and feedback.

I also head up our successor ARC-funded research project on longevity risk. It has a wider scope, and as part of that project, we are looking at risk management, and index-based hedging is part of what we are looking at. Much of the content of this paper, but also the discussion that follows, will be potentially influential in terms of the particular directions that we take within our own research.

I would like to make some comments with occasional questions. By way of introduction, we have seen the focus here on index-based hedges. I see these, although there has not been very much activity on index-based hedging so far, as being becoming increasingly relevant and prevalent in the coming years. Otherwise, if it does not happen in that sort of way, it is going to be rather difficult to satisfy the demand that is coming from the hedgers, be it the pension funds or the insurance companies.

Part of that supply has to come from the capital markets. This might not be achieved necessarily in individual steps. The role of index-based hedges might not be so much for pension plans to transfer to insurers or reinsurers, but it might be part of a risk-transfer chain. What we might see is that there are customised swaps between pension plans and insurers. That comes out of the sampling risk that has been an important part of this paper, where we see that for the smaller pension plans – probably most pension plans in terms of the numbers being talked about – the sampling risk is perhaps too big to be thinking about an index-based hedge.

They might have a customised swap with a reinsurer. Then the reinsurer can release capacity by itself by issuing some sort of index-based hedge onto the market. This is a way where I see that this work might be particularly useful.

Moving on now to talk about the paper. The first part is the data part of the paper. It is certainly very useful in terms of having the different datasets for road testing different parts of the technology and to see what the differences are.

The second part introduces the two-population models. The third part introduces the risk measures and hedge effectiveness. Fourth, there are hedging instruments. Then fifth, the extensive part, the core of the paper, is the analysis; a lot of numerical tests on different models and many variations.

On the data side, there is not much to say except that in terms of what is in the paper and what is not in the paper. What is missing perhaps would be a summary of each of the datasets in terms of numbers of exposures. That then feeds into the later parts of the paper in terms of trying to interpret some of the numbers that are coming out.

There is also a question of whether the data might or might not be made available because it is certainly quite useful for us, researchers and people in industry, to be able to try out our own models with various datasets to see whether they are going to work.

Then another question is when it comes to the model calibrations: when they did the model calibrations, fitting the time series models, et cetera, did the authors find any volatility bias creeping in for the smaller population? For the smaller population was there increased volatility which was highlighted in the phase 1 paper?

Moving on to the part on risk measures. That was well introduced in the presentation. The four measures were later reduced to three. What was not completely clear in the paper was, when you were looking at value at risk in particular, was it value at risk based on the run off distribution or was it 1 year ahead value at risk, which is embedded in Solvency II?

Obviously, insurers, particularly in the UK, will be more focused on the 1-year value at risk. But then there are related issues in terms of having to recalibrate the model after 1 year as well. So there are differences in terms of which value at risk it is that you are measuring. And certainly hedge effectiveness calculations on that 1 year would be welcome. It might be that what is in the paper is 1 year but I think it is the run-off.

Moving on to hedging instruments, the authors have considered a range of what I would consider to be plain vanilla types of contracts: swaps, and so on.

In the current paper, there is a longevity swap that forms an initial core part of the analysis. The floating-like payments are based on the national population without any adjustment. Something that could be easily incorporated into that example is to include experience ratios, which would adjust the national population mortality rates year-on-year to level them up or down so that they are seen from the outset as having the same expected values rather than one of them being 50% throughout.

Then in the numerical examples, you see, particularly when you look at the index of multiple deprivation groups, as discussed in the paper, group 3 in the middle does best. That is merely because the mortality rates for group 3 are very similar in terms of their levels to the national population, or at least the expected levels, whereas the high deprivation people die off very quickly compared to the national population, and the low deprivation groups die off much more slowly. So you get a mismatch later on.

Much of that could be addressed through introducing experience ratios and therefore you may well get comparable levels of hedge effectiveness by incorporating those factors. These are factors which people in the industry are already using with these sorts of contracts.

For further work, optionality could be introduced into some of these contracts. For example, in the Aegon deal in the Netherlands that was based on national mortality, it included a cap and a floor on the payouts. That helps to move on to the capital markets. What they want are not just time limited but also limited liability in terms of the payment.

The paper has a very good account of the three types of basis risk, which are population basis risk, sampling basis risk and then the structural basis risk. The impact of sampling risk is well covered in the paper here in terms of the different scheme sizes. There are many examples where you can see what the impact is of having 10,000 or 1,000 members.

When it then goes on looking at population basis risk and structural basis risk, the reader has to work a little bit harder to separate those two things out. Perhaps one way ahead in terms of getting a

better grip on the impact of population basis risk versus structural basis risk is to re-run the simulation but to include perfect correlation rather than less than perfect correlation.

In that way you would filter out the population basis risk and what would be left would be the structural basis risk. That is something which is worth trying. Part of the reason for that is that it is important for regulators to be able to know something about the population basis risk in particular and its impact.

On the optimisation side in the paper, which is touched on a little, the authors are typically optimising in the sense of maximising the hedge effectiveness. In many circumstances, it might be more complex than that. For example, there is going to be a trade-off. You would be wanting to think about the risk appetite of the hedger, and then you also have to think about what the price of the hedge is that is going to be put in place. If it is more expensive, then you might still want to hedge but not at the same level as if it were a lower price.

So there is an interesting bit of work still to be done in that direction to take what is in this paper through to a question of perhaps something like what the economic value added or destroyed would be as a result of putting a hedge into place. That all comes down to questions of risk appetite and economic capital calculations, as well as regulatory capital calculations.

Sections 4 and 5 contain very helpful and very extensive numerical results. The results confirm a lot of intuition. The middle group for the index of multiple deprivation does better than the other groups because their trajectory is similar to the national population rather than dying off faster or slower. That is one of the intuitive results. You can see for the more extreme groups that you get less or a lower hedge effectiveness.

Also intuitive were the results on schemes size and interest rates. What is new and very useful in the paper is that while we know in what direction these things are going, what is the magnitude of these impacts? There is also the magnitude of the variations between the different non-Index of Multiple Deprivation (IMD) groups, the Australian pensioners and also the CMI data sets.

Again, there could still be more interpretation of the differences. Are the differences for these different groups because of the duration of the average age or average expected lifetimes, or is it owing to the volatility of the book population relative to the reference population? Or is it perhaps due to differences in the intrinsic correlation? It is going to be a combination of these factors. When you are analysing one data set, you want to be thinking why this population is different from the middle IMD group.

Lastly, model risk is also an important part of the paper, an important contribution that comes out of it, and this is a section which should be of interest to regulators. In the UK, we have the Prudential Regulatory Authority. They have specific guidelines for when you are doing single population modelling and talk about the combined use of several single population stochastic models. Over time, it may be that the regulators might wish to extend the single population guidelines to two populations or even more.

Dr Li (responding): Regarding the last point on model risk, we realise that using different models can sometimes come up with quite different numbers.

Apart from simply applying different models, generating different results and comparing the results, one useful alternative, at least academically, is to use Bayesian methods to try to allow for the model risk. But using Bayesian methods would be much more time-consuming and computationally demanding.

Regarding the point about the cost of the hedge, we agree that it is certainly a next step to do. In the report, we assume that the forward swap rate is the expected value of the future rate. It implies a zero-risk premium, which is a convenient assumption for this project. When there are more data, particularly market price data of mortality- or longevity-linked securities, we hope to do further work on this area and try to price how much an index-based swap would cost, and then see whether it is worth taking this to implement a hedge.

Another point we find very helpful is your suggestion that we can assume a perfect relationship between the two populations and filter out the demographic basis risk and then focus on the structural basis risk. We certainly want to try that.

Lastly, one interesting thing you mentioned was the example of the index swap by Deutsche Bank in 2012, a 12 billion euro index-based swap targeted to the capital market investors. They set it to only 10 years, which can be seen as the market perception of the longest period that the investors are willing to take the risk. They have caps and floors. It would be very interesting to conduct further research on these caps and floors, other option features, and other derivatives and see the resulting hedge effectiveness.

Dr C. I. Tan (responding): I focus my response on a couple of things regarding the model calibration. Because of the limited size of the book data, when we tried to fit the dataset there was very large sampling variability at times. When we moved onto the time series parameter estimation, we found that a small book size and a short history would result in serious estimation problems.

So the volatility is a key concern. If we want to do this analysis in a more informative way, we not only need a bigger data size but also need a longer history, which is also highlighted in the phase 1 decision tree framework. Our findings doubly confirmed that.

Regarding the value-at-risk, it is not the 1-year value-at-risk, it is the run-off value at risk. It will be interesting to extend the analysis and look at the 1-year value-at-risk and see whether the results might be robust or might change.

Regarding the comment on the experience ratio and the optionality of the hedging instrument, these are very interesting suggestions. For future research, we could look at how the different pay off structures of hedging instruments with different option features or experience ratios might impact, increase or reduce the longevity risk reduction.

Regarding the comment about dissecting the population basis risk and structural basis risk, because of how the three basis risks are defined, sampling basis risk can be stand-alone, and perceived as the risk associated with the number of lives of the pension plan.

Another way to look at the sampling risk is on the data set sampling risk from fitting the models to the data.

Regarding the population risk and the structural risk, if we go back to a single population, with the same reference population and book population, then there is no demographic basis risk and we can quantify the structural basis risk. While pension plans are paying the liabilities in terms of survival probability, using, say, q -forward, which is in terms of death probability, would lead to a mismatch and so structural basis risk. If we extend this scenario to different reference and book populations, then the resulting structural basis risk contain the features of both demographic and structural basis risks. In that sense, the perfect correlation analysis is a very interesting suggestion. We will definitely look at that, as the population and structural basis risks can be entangled.

Regarding the optimisation objective, in our report, we concentrated on the risk reduction. It is certainly very informative to look at how we combine the risk appetite and the price, because in the phase 1 sessional meeting, one of the questions raised was about cost and benefit analysis. So the pricing would certainly be an important area for future research.

Mr S. D. Baxter, F.I.A.: First, congratulations to the authors on a fantastic paper that has added much to our knowledge of the sensitivities and the key considerations when it comes to hedging of longevity basis risk.

I was struck by a number of important additions that could be added to the sensitivities. If I understand the paper correctly, throughout your theoretical portfolios you are looking at paying an annuity of £1 (or other such fixed amount) per policyholder/pension scheme member.

In reality, the portfolios that we deal with as an industry have considerable variation in annuity payments between individuals. This serves to simply reduce the effective size of the book in terms of the sampling risk and probably reduces the hedge effectiveness compared to that presented.

I also suspect that once you bring in the practical aspects of the instruments available in the market the analysis changes. Of particular concern is the limited duration of such contracts, resulting in a settlement payment at the end of the term. This payment is typically based upon a formulaic extrapolation of longevity beyond the term of the contract that is agreed upfront. The extrapolation potentially reduces the hedge effectiveness of the contract compared to the situation where the contract is held in perpetuity.

However, it strikes me there are some simplifications being made that might also be reducing the index-based hedging benefits presented.

First, I suspect the analysis relates to a buy and hold strategy. You buy an instrument at the outset and simply hold it throughout the run-off without modification. If that is the case, then it ignores the option of restructuring your hedge over time. Such rebalancing may mean that you get better overall hedge effectiveness, although there are some frictional costs to allow for here.

Second, a possible way of enhancing the hedge effectiveness is to have a small number of sub-population indices. A key question is: would that improve the hedge effectiveness or would the benefit be marginal? For instance, if you model longevity patterns at the deprivation decile level, and consider a situation whereby the longevity swap is linked to outcomes at the deprivation quintile level, what is the hedge effectiveness?

Finally, I question the suggestion made that the early adopters of such instruments – for instance, insurance companies and reinsurance companies – will inevitably use them to free up balance sheet capacity. They have quite sophisticated balance sheets and metrics that are considerably more complex than the natural ones that you have chosen. For example, insurers will wish to see efficacy on 1-year based Solvency II metrics, and more broadly the interactions across the corporate balance sheet. Have the authors given any thought to, or have any plans for looking at the more holistic picture and understanding some of the dynamics of practical interest to insurers and regulators?

Dr G. Coughlan: I also congratulate the authors on a fine paper clearly presented here.

When I consider basis risk, the important thing is hedge effectiveness. Hedge effectiveness you can view in two ways, which are different conceptually, practically and economically. That is hedge effectiveness in the body of the distribution versus hedge effectiveness in the tail, so what you might have referred to as your volatility versus parametric.

The hedge effectiveness in the body of the distribution was susceptible to noise that may come from sampling error; it may come from modelling error; it may come from the intrinsic variability in the mortality processes underlying the two populations. But when you get into the tails, and if you get into the extreme tails, that becomes less important. It is also the most economically relevant area.

The difficulty with any kind of modelling of hedge effectiveness in those extreme areas is coming up with a realistic simulation model appropriately correlated with extreme scenario movements, i.e. when the mortality outcomes are far from the expected path.

Do you have any insights into how we can make some progress on this kind of simulation model capturing more realistically the extreme scenarios? That is what hedges are most concerned about.

The conceptual linkage there is, particularly if one population is a sub-population of the other, that the bigger movement, the greater you would expect the pull of the coherence referred to get you to higher hedge effectiveness in the future.

Dr Tan (responding): Regarding the use of dynamic hedging raised by Steven [Baxter], in this report we consider a simple case of single cohort hedging and make use of a cohort-specific longevity swap with zero risk premium. We keep track of the cohort over time. There is no need to change the position once you have established a position at the start. A good alternative would be using q-forwards that track the mortality movements over different time periods.

So when we are talking about a single cohort, we may use, say, four to five q-forwards catering for different ages. This is especially important when we move to multiple cohorts. The multiple cohort plan analysis can be found in the report.

We also demonstrate the use of multiple longevity swaps to cater for different cohorts. If you have a range of cohorts, obviously a limited number of longevity swaps may not be sufficient.

Regarding the point about the hedging strategy at the outset, because of simplifying assumptions, there is no need to adjust the position over time.

If the hedging costs are taken into account and we want to adjust the position, that can be done. In that case, probably a longevity swap might be suitable. One can first initiate a swap position for the

next 10 years. Then 10 years later, the plan can change the position. In that sense, we are doing dynamic hedging. In theory, the hedge effectiveness will be better, given that the costs are not too high. But the extent of improvement in practice would need to be further investigated.

Regarding the point by Guy [Coughlan] on the tail, that is a very difficult question. In the existing literature, the focus is often on the variance and the resulting optimal hedge ratio.

With longevity swaps, when it comes to optimising the variance or the standard deviation risk reduction, that can be done easily.

When it comes to tail measures, it is difficult. When we calculate the tail measure, we are looking at the most extreme 0.5% value-at-risk or expected shortfall. In the estimation process, we need to rank the simulated scenarios before the hedge and after the hedge. Because of the way of the optimal weight is determined, a particular scenario could be ranked very differently before and after the hedge. As far as we are aware, there is no common mathematical measure or property for this purpose.

If we can tackle the problem of re-ranking, using some mathematical measure that we are not aware of yet, then we can do some more meaningful analysis on the tail risk hedging problem. At the moment, we are not sure what a good indication of potential tail risk reduction is.

Let me go back to the longevity swap with regard to variance reduction as a comparison. In fact, what we need to do is just to simulate 10,000, 5,000 or 1,000 scenarios and create a longevity hedge based on the variance and co-variance, because the optimal hedge ratio is actually a function of the variance and covariance. There is no need to rely solely on the numerical optimisation.

But in this project, we are looking at the 99.5% value-at-risk. That is why at the moment numerical optimisation is the best alternative. If we can solve the re-ranking problem, especially at the tail part, then that would be useful for industry practitioners.

Dr Li (responding): A quick comment on extreme scenarios. Extreme scenarios are extreme after all. It is difficult to find data to allow for it. But in our report we have added something on allowing for potential extremes. In the solvency capital requirement under Solvency II, one way to allow for the so-called longevity shock is to decrease all mortality rates by 10%.

So we tried, subjectively, to decrease the book and the reference mortality rates by 10%, 20%, 30%, by different amounts between the two sets of rates, and see the effects. It turns out that the hedging effects are not bad. The hedging results are indicating that at least a third of the loss could be avoided by having an index-based hedge, when we allow for those extreme movements in mortality rates subjectively.

We also tested incorporating mortality jumps and structural changes stochastically. Those are more extreme changes. After we allowed for these effects, the overall variability increased a lot, and the tail behaviour became different. Nevertheless, the simulated hedging results turn out to still appear quite well.

More details can be found in our report. Right now, the way we have set it up is more subjective because we do not have data for the tail. There were a few wars and epidemics in the past, and there

are little data on these extreme events. So far we have tried arbitrary allowance for that and the simulated hedging results are still broadly in line with the others.

For future research, different models and different scenario testings can be done more specifically for the tail extremes.

Mr S. Rimmer, F.I.A.: In Figure 8 where you had hedge effectiveness in terms of a series of factors, it seemed like some of the factors were decisions about the modelling that had been done to understand the contract.

Could you explain more about how that feeds through into an increased or decreased hedge effectiveness? The picture that I have in my mind is we have two reinsurers who have a 50% quota share on the same portfolio, the same contract. Is there some sense to which the way that they are doing their internal modelling and perceiving the contract affects hedge effectiveness?

A factor in how much risk reduction you get seems to be a choice of our modelling. I am wondering whether we can understand a bit more. Am I misunderstanding what we mean by risk reduction or is there some extent to which the way you perceive the risk is what we are measuring?

Dr Tan (responding): If I interpret it correctly, your question is whether the risk reduction that we are getting is a matter of the assumption settings, the hedging environment, and so on?

These results are based on the hundreds of simulations that we have done. It does not directly suggest that using a certain assumption might lead to a better or worse risk reduction, but rather it is a manifestation of different sensitive or insensitive assumptions that might give rise to different simulated hedging results.

For example, if the book size is rather small, then, according to our numerical results, the risk reduction, most of the time, would be below 50%.

I would say the first five factors are important, because we categorise the first five factors as something we could not control to a certain extent, and the last five as something that we could control, for example, using a different model or using a different simulation method.

With regard to the first five factors, we are taking what is given, and we estimate the risk reduction. I agree that we have an option to play around with different methods and different variables when it comes to the last five factors. In this project, apart from the data size and the demographic structure, the rest of them are based on pre-specified assumptions.

The time series process is a very significant part – using a different time series process might lead to good or poor simulated hedging results.

What we find is that the simulated future variability of the book and reference populations is the most critical consideration. If we are using a time series process that supports or implies a bound on variability, then we are going to get a good risk reduction estimate.

But again when it comes to choosing an appropriate time series process, we have to exercise our own judgement whether a coherent or whether the bounded or unbounded variability might be the case in the future.

The Chairman: If this research was to go to phase 3, what do you think would be the most productive areas to investigate, and what do you think the challenges of such an investigation would be?

Dr Li (responding): Based on the outcomes I have seen in phase 1, and based on the results in phase 2, we have already stretched most of the things we could do, based on the data that we have collected. If there is going to be a phase 3, we need to collect more data to do the work, especially the current book data available limits the things that we could have done.

Apart from getting more data and trying more models, the next thing is to communicate the results with practitioners and regulators to see whether more practical elements can be put into the research and how this index-based hedging can be translated into reality, including possible capital savings, trading index-based securities, and market pricing of such securities.

Dr Tan: Phase 3 would be going back to the question if a pension plan is given a choice between a bespoke de-risking solution and an index-based longevity swap or mortality derivative, what are the costs and the benefits of each choice?

For example, would the regulator be convinced that the capital is adequate for that purpose if an index-based solution is used? It is about further communication between the pension plans, the regulators, all the stakeholders on the demand side to come together and consider what can be done and what is further needed.

On the supply side, communicating the advantages of investing in a longevity asset class to market players that do not have mortality or longevity exposures is something to consider.

I think the main challenge is on the demand side from a practical point of view.

The Chairman: Andrew (Cairns), your comment about the role of index-based transactions as part of a chain of transactions of risk transfer – I had not actually heard it phrased that way before – conceptually that seems to make sense to have the bespoke transactions to meet the needs of the trustees' pension schemes.

When you are dealing with professional reinsurers probably to the capital markets, the index-based solutions do seem a natural way. We keep expecting this market to develop – it depends how you count it – but beyond 10 billion or 15 billion per year that it seems to have been running along.

There remains quite a lot of areas that can be investigated. The need for more data remains an open one. There are obviously very large pension schemes, particularly state pension schemes in countries outside the UK, elsewhere in Europe and elsewhere in Asia. There is a number of datasets that have scale but maybe not the history.

European datasets, maybe, have a longer history but tend to be smaller in scale than potentially some of those elsewhere. That certainly is an area for investigation going forward.

I express my, and all here, thanks to the presenters of the paper.